Middlesex University London

Middlesex University Research Repository

An open access repository of

Middlesex University research

http://eprints.mdx.ac.uk

Mohammed, Ameer, Zamani, Majid, Bayford, Richard ORCID logoORCID: https://orcid.org/0000-0001-8863-6385 and Demosthenous, Andreas (2017) Toward on-demand deep brain stimulation using online Parkinson's disease prediction driven by dynamic detection. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25 (12) . pp. 2441-2452. ISSN 1534-4320 [Article] (doi:10.1109/TNSRE.2017.2722986)

Final accepted version (with author's formatting)

This version is available at: https://eprints.mdx.ac.uk/21775/

Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners unless otherwise stated. The work is supplied on the understanding that any use for commercial gain is strictly forbidden. A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge.

Works, including theses and research projects, may not be reproduced in any format or medium, or extensive quotations taken from them, or their content changed in any way, without first obtaining permission in writing from the copyright holder(s). They may not be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s).

Full bibliographic details must be given when referring to, or quoting from full items including the author's name, the title of the work, publication details where relevant (place, publisher, date), pagination, and for theses or dissertations the awarding institution, the degree type awarded, and the date of the award.

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:

eprints@mdx.ac.uk

The item will be removed from the repository while any claim is being investigated.

See also repository copyright: re-use policy: http://eprints.mdx.ac.uk/policies.html#copy

Towards On-Demand Deep Brain Stimulation Using Online Parkinson's Disease Prediction Driven by Dynamic Detection

Ameer Mohammed, *Student Member, IEEE,* Majid Zamani, *Member, IEEE,*
Richard Bayford, *Member IEEE,* and Andreas Demosthenous, *Senior Member, IEEE*

***Abstract*—In Parkinson's disease (PD), on-demand deep brain stimulation (DBS) is required so that stimulation is regulated to reduce side effects resulting from continuous stimulation and PD exacerbation due to untimely stimulation. Also, the progressive nature of PD necessitates the use of dynamic detection schemes that can track the nonlinearities in PD. This paper proposes the use of dynamic feature extraction feature extraction and dynamic pattern classification to achieve dynamic PD detection taking into account the demand for high accuracy, low computation and real-time detection. The dynamic feature extraction and dynamic pattern classification are selected by evaluating a subset of feature extraction, dimensionality reduction and classification algorithms that have been used in brain machine interfaces. A novel dimensionality reduction technique, the maximum ratio method (MRM) is proposed, which provides the most efficient performance. In terms of accuracy and complexity for hardware implementation, a combination having discrete wavelet transform for feature extraction, MRM for dimensionality reduction and dynamic k-nearest neighbor for classification was chosen as the most efficient. It achieves mean accuracy measures of classification accuracy 99.29%, F1-score of 97.90% and a choice probability of 99.86%.**

***Index Terms*—Biomedical signal processing, deep brain stimulation (DBS), dimensionality reduction, dynamic detection, dynamic pattern classification, feature extraction, Parkinson's disease, semi-synthetic LFP generation.**

I. Introduction

CLINICAL deep brain stimulation (DBS) for Parkinson's disease (PD) uses continuous, high frequency voltage or current pulses in order to mitigate PD. The major setbacks of present clinical DBS are stimulation induced side effects and shortening of pacemaker battery life [1], which can be addressed using on-demand DBS. It regulates stimulation by controlling the stimulation intensity and timing using feedback signals from the stimulation site [2]. On-demand DBS has the potential of improving stimulation efficacy, reducing power consumption and reducing side effects [1].

Several studies have proposed different ways of obtaining minimally invasive feedback signals for on-demand DBS. Primary focus has been placed on internal, e.g. LFP (local field potentials), ECoG (electro-corticography) [1], [2]; and external (e.g. EMG) neuro-electrophysiological signals [3]; biochemical signals [4]; and mechanical signals [3]; some of which have shown promising results. Electrophysiological signals obtained from external measurements are less suitable for detecting spontaneous PD activity due to their low correlation with rigidity and bradykinesia [5], [6]. The discomfort experienced by patients due to externally attached sensors may also be an issue. Using biochemical signals, detection may not be instantaneous, and may be hindered by the need for bulky and complex devices in some cases. Internal electrophysiological signals such as LFP are known to provide an indication of PD symptoms such as rigidity, slowness of movement or tremor [7]. Also, recordings can be obtained from the same electrodes that are used for stimulation [8], making them minimally invasive. However, LFP correlations to bradykinesia and rigidity are functionally different from those of tremor [9]. This reinforces the need to identify distinguishing features in the data for individual patients.

Various studies have mainly focused on monitoring beta band LFP only [1], [10], [11]. However, using only beta band LFP may not be sufficient, as they have not displayed satisfactory consistency across time and patients [1], [12]. Furthermore, the correlation of gamma [7], [13], [14], and tremor [15] bands with PD symptoms, raises more questions on the suitability of using beta band activities alone. In essence, triggering DBS using characteristics from only a single band may be ineffective. Alternatively, in order to create robust (effective) feedback algorithms, identifying the most relevant recording channels (in multichannel recording) and frequency bands can provide a better mapping between LFP recordings and disease states. Hence, for every patient and at certain intervals of time, the channels and/or frequency bands that display the most pronounced variation between PD and non-PD events can be determined and used to detect PD states; this implies using fewer features, which are dynamically updated.

The mapping between the LFP features and PD states may not be straightforward, which can cause the selected features to perform poorly. Detection accuracy can be improved using

Manuscript received May 28, 2016; revised November 09, 2016 and May 19, 2017; accepted June 23, 2017. This work was partially supported by the Presidential Special Scholarship Scheme for Innovation and Development (PRESSID), Nigeria.

A. Mohammed, M. Zamani and A. Demosthenous are with the Department of Electronic and Electrical Engineering, University College London, WC1E 7JE London, U.K. (e-mail: ameer.mohammed.13@ucl.ac.uk; m.zamani@ucl.ac.uk; a.demosthenous@ucl.ac.uk).

R. Bayford is with the Department of Natural Sciences, Middlesex University, The Burroughs, London NW4 6BT, U.K. (e-mail: r.bayford@mdx.ac.uk).

an additional dynamic stage consisting of dynamic classifiers. The processing chain is shown in Fig. 1. The implementation of a robust PD detection scheme is necessary because inaccurate detection results in administering stimulation when it is not required, and this may lead to stimulation induced side effects [16]. Or, inaccurate detection may result in the non-administering of stimulation when it is required, which may worsen patient condition [17].

This paper examines and evaluates a subset of brain machine interface algorithms suitable for on-chip implementation of PD detection in real time with high performance and low complexity. Computationally efficient on-chip (online) PD detection would facilitate the development of fully implantable closed loop DBS systems that could automatically adjust stimulation parameters by the brain response in real time. The optimum combination of algorithms in terms of detection accuracy and computational complexity consisting of feature extraction (FE), dimensionality reduction (DR) and dynamic classifier algorithms is identified. A novel DR technique, the maximum ratio method (MRM) is proposed. The algorithms are tested using representative PD and non-PD datasets to choose the best algorithm combination for real time hardware implementation. This paper presents a thorough study of the preliminary work in [18].

The rest of the paper is organized as follows. Section II details the methodology for evaluating the algorithms. Section III describes the candidate FE, DR and dynamic classifier algorithms. Section IV compares the performance of the examined algorithms in terms of accuracy and complexity. Discussion and concluding remarks are presented in Section V and Section VI respectively.

II. Methods

A. LFP Datasets

In order to evaluate the performance of the algorithms, accurately labeled LFP recordings for PD patients are needed. Studies have used modelled subthalamic nucleus (STN) LFP recordings for controlling DBS [10], [19]. These may be inadequate since LFP recordings have been reported to be modulated by cognitive, emotional and behavioral tasks [7], [14], [20] which introduce unpredictable variations. Using real physiological recordings produces better models that incorporate the dynamic variations present in LFP signals. So far, studies in PD detection have been hindered by the unavailability of standard databases of Parkinsonian LFP signals that could be used for evaluating algorithms. A possible solution is to use LFP recordings from PD patients subjected to levodopa (L-dopa), a common pharmacological therapy for PD patients, consisting of periods in which patients are ON and OFF L-dopa. The ON L-dopa periods are periods when L-dopa medication is effective and is normally accompanied by little or no PD symptoms, while OFF L-dopa periods are when PD symptoms return, signifying periods of motor deficit [21], [22]. The test datasets used were LFP recordings from the STN of subjects exhibiting a combination

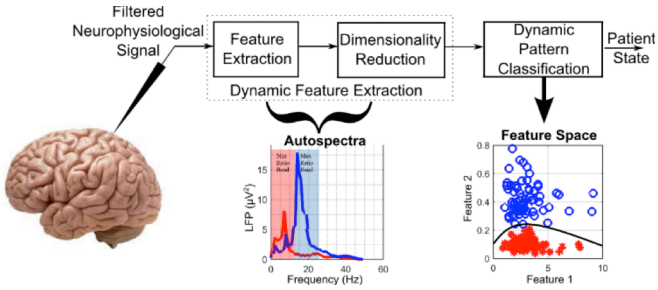

Fig. 1. Typical bio-signal processing chain of PD state determination during on-demand DBS.

of bradykinesia and/or rigidity during the onset of PD, with less noticeable tremor. Recordings were made from nine patients with PD who had bilaterally implanted DBS electrodes in their STN and are referred to as dataset A–I. Each patient recording contained separate ON and OFF L-dopa data between 5 to 10 minutes long.[1] ON and OFF L-dopa LFP data are referred to as non-PD and PD data respectively. In order to increase the length of recordings which mimic the unpredictable nature of LFP recordings, semi-synthetic datasets can be modelled using the LFP recordings from each dataset. There are a number of approaches to achieving this as summarized below.

1) Semi-Synthetic Data Generation Techniques: Semi-synthetic data generation provides the flexibility to manipulate the signal characteristics such that all underlying conditions are represented. This enables conclusions that could be extended to an entire population. LFP are extracellular activities obtained from a localized population of neurons, making it necessary to utilize inherent statistical properties that maintain the information modelled by individual samples, as well as segments of the signal. Nevertheless, there are a number of ways this could be done depending on the property to be exploited. Two statistical properties that could be exploited are non-Gaussianity and stationarity of the signals.

- For non-Gaussian signals, the pattern of activity from statistically independent sources that contribute to the parent signal can be obtained using spatial source separation techniques. Contributions from various spatial sources can be modelled to create semi-synthetic data sources. This is pertinent because the brain models information in a statistically sophisticated way using multi-physiological activity [23]. However, most neural signals are Gaussian [24], making it

[1] The data was obtained from the Department of Clinical Neurology, University of Oxford. Recordings were made prior to the connection of a subcutaneous DBS pacemaker and stimulation was completely off during recording. Details on the daily drug dosage, on and off unified Parkinson's disease rating scale (UPDRS) score and dominant symptoms for eight of the nine patients are summarized in [1]. The permanent quadri-polar macro-electrode used was model 3389 (Medtronic Neurologic Division, Minneapolis, MN) consisting of 4 platinum-iridium cylindrical contacts. Its contacts are numbered 0, 1, 2 and 3, with 0 being the most caudal and 3 being the most cranial for both right and left electrodes – making a total of eight monopolar channels for each patient.

difficult to separate them into their independent sources.

- Alternatively, stationarity could be exploited. For stationary or weakly stationary signals, autoregressive moving average (ARMA) models could be fitted to the signal [25].
- Finally, for non-stationary signals, an autoregressive integrated moving average model (ARIMA) can be used.

Since the original LFP recordings in this study are weak (or wide) sense stationary (WSS), an ARMA model was fitted.

2) ARMA Model: For each channel in each dataset a suitable ARMA (p, q) model was fitted to the original LFP signal – for both PD and non-PD data. An ARMA (p, q) model has autoregressive (AR) process of order p, and a moving-average (MA) process of order q [26]. As an example, an ARMA (1, 1) can be represented as,

$$X_t - \mu = \alpha_1(X_{t-1} - \mu) + Z_t + \beta_1 Z_{t-1} \qquad (1)$$

where the observations are $X_1, X_2, \ldots, X_K$, with X_t as the current observation and X_{t-1} as the previous observation; with μ as the mean, α_1 and β_1 are the AR and MA parameters respectively; while Z_t is a purely random process with mean zero and variance σ_Z^2. For model fitting, the original LFP signal was low-pass filtered (tenth-order Butterworth filter, with cutoff frequency at 50 Hz which is in the range of PD relevant LFP [14]), down-sampled to $f_s = 128$ Hz (from 2048 Hz) and detrended. Detrending makes the datasets stationary. In selecting the appropriate order and parameters for the AR and MA terms, residual analysis was used (residual = observation – fitted values). For a good model, more than 95% of the residual autocorrelations should lie within the range $\pm 2/\sqrt{K}$, where K is the number of observations. After fitting the appropriate models for both PD and non-PD data, longer recordings consisting of PD and non-PD periods are created. The process is summarized as follows:

1. The PD data is firstly divided into one second epochs. A 300 second PD recording thus consists of 300 epochs. These epochs are used as templates for the semi-synthetic data generation.
2. All the epochs are segmented into three regions based on their similarity – measured using the normalized cross-correlation between epochs. Thus, a 300 epoch dataset is divided into three segments consisting of approximately 100 epochs (templates) per segment. The aim of this segmentation was to ensure that any three succeeding PD epochs attached to create long PD periods were selected from different segments, which introduces a distinctly different characteristic.
3. Using the fitted ARMA models, inferred residuals and individual epochs, a number of possible forecasts can be realized. Using an original epoch, Fig. 2(a) shows forecasts from two possible Monte Carlo paths taken by a semi-synthetic epoch. For each one second original epoch, 100 Monte Carlo forecasts (with one second duration) are made. Thus, for LFP recordings consisting of 300 templates, 30,000 semi-synthetic epochs were created.
4. Steps 1 – 3 are repeated for the non-PD data.
5. To create long recordings consisting of alternating PD and non-PD episodes that imitate the progression of LFP signals in PD; a Poisson distribution was used to define the duration of the PD and non-PD episodes. This was done to observe how well the algorithm performed when tested with few and rampant PD episodes that mimic real life situations [27], [28]. The original and newly generated epochs in steps 1–4 were used to synthesize twelve-hour long LFP recordings.

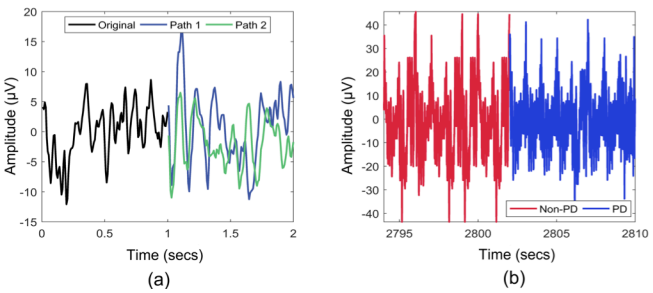

Fig. 2. Semi-synthetic LFP generation. (a) An original LFP epoch and two of its possible forecasted paths based on Monte Carlo simulations. (b) Snapshot of semi-synthetic LFP recordings consisting of PD (blue) and non-PD (red) segments. Both figures are for patient (dataset) C.

During LFP synthesis, random permutation sampling was used for epoch selection. To maintain all frequency components present in the original signal and to avoid introducing unwanted frequencies, successively attached epochs were slightly overlapped and averaged at the overlapping points [29]. Fig. 2(b) shows a sample of the synthetically generated LFP recordings indicating PD (blue) and non-PD (red) periods. Portions of the synthetic LFP were used to train the system to detect the patient state for unseen recordings. During FE, in instances where the sliding window selects segments of LFP data that have nearly equal proportions of non-PD and PD, are tagged as transition states.

B. Complexity Estimation

Complexity is quantified using a weighted cost of the number of operations (NOP) and the estimated microchip area. The smallest unit for the NOP is 1-bit addition. Subtraction is considered to be equal to addition. Each multiplication or division is considered to be L additions, using a quantization of L bits/sample. Microchip area is divided into logic and memory area. Adders, subtractors and comparators are assumed to be the same size. Following the procedure in [30] for a 90 nm CMOS process, each 1-bit adder is estimated to be 20.46 μm^2. The size of multipliers and dividers are assumed L-times the size of adders for L bits/sample quantization. Memory is calculated based on the number of registers needed for each computation. A 1-bit register size was reported to be 15 μm^2 for a 90 nm CMOS process [30]. By labelling each combination of detection stages as n, and the complexity cost of each combination as CompCost$_n$, the weighted complexity cost can be calculated as,

TABLE I
METRICS FOR EVALUATING ALGORITHMS

Error Cost	Complexity Cost	Combined Cost
F1-score	NOP	Error Cost
Choice Probability	Area	Complexity Cost

TABLE II
THE VARIOUS ALGORITHMS TO BE EVALUATED

Dynamic Feature Extraction		Dynamic Pattern Classifiers
Feature Extraction (FE)	Dimensionality Reduction (DR)	
DWT	MRM	Dynamic k-NN
STFT	PCA	Dynamic LR
		Dynamic SVM

$$\mathrm{CompCost}_n = \frac{1}{2}\left(\frac{\mathrm{NOP_n}}{\max(\mathrm{NOP}_1,\ \mathrm{NOP}_2, \ldots \mathrm{NOP}_{n-1},\ \mathrm{NOP}_n)} + \frac{\mathrm{area}_n}{\max(\mathrm{area}_1,\ \mathrm{area}_2, \ldots \mathrm{area}_{n-1},\ \mathrm{area}_n)}\right) \quad (2)$$

where NOP_n is the NOP for the combination labeled as n, and max(…) computes the maximum value of the various combinations. The complexity cost assigns a 50% weight to both NOP and area. A maximum complexity cost of 1 can be obtained for a combination concurrently having the largest area and largest number of operations. Table I summarizes the various costs used in evaluating the algorithms.

C. Detection Accuracy Calculations

The PD detection scheme consisting of FE, DR and dynamic pattern classification, were tested in MATLAB. Combinations consisting of all the three stages in the PD detection scheme were evaluated using the algorithms in Table II. Each combination was evaluated using the following metrics:

1) Classification Error (Err): It gives the proportion of incorrectly classified test cases and is defined as

$$\mathrm{Err} = \frac{\mathrm{FN} + \mathrm{FP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}} \quad (3)$$

where TP represents the true positives (the epochs that are actually PD and were correctly detected to be PD), TN represents the true negatives (the non-PD epochs that were correctly detected), FN represents the false negatives (the PD epochs that were wrongly detected to be non-PD) and FP represents the false positives (the non-PD epochs that were wrongly detected to be PD).

2) F1-score: It is used in situations like PD detection, where a single metric is needed to analyze the performance of a classifier in terms of precision and sensitivity [31]. The F1-score is defined as

$$\mathrm{F}_1 = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}. \quad (4)$$

3) Receiver Operating Characteristics (ROC): ROC is a plot of sensitivity and false positive rate and has an area under the curve (AUC) of between 0 and 1. It is used to evaluate the performance of the various detection algorithms. The AUC of the ROC, also called the choice probability (CP), represents the probability that the detector will correctly classify an event in a two-alternative forced-choice classification. See [32] for a more detailed description.

4) Error Cost: Each complexity cost ($\mathrm{CompCost_n}$), has a corresponding error cost ($\mathrm{ErrCost_n}$) that is calculated at 10% classification error and is

$$\mathrm{ErrCost}_n = \frac{(1 - F_1) + (1 - CP)}{2}. \quad (5)$$

A maximum error cost of 1 can be obtained for a combination having $F_1 = 0$ and CP = 0.

D. Model Assumptions

The algorithms evaluated were chosen based on their efficiency in previous brain machine interface and bio-signal processing applications. The process involved evaluating a subset of FE, DR and machine learning models used in pattern classification. The detection scheme uses ON and OFF L-dopa signals as representative data for non-PD and PD cases respectively. It is also assumed that training occurs only once a day.

III. ALGORITHMS

A. Feature Extraction (FE)

In order to extract meaningful information, acquired physiological data which is normally in the time domain, is transformed to a computationally efficient form for further processing. In applications like PD state detection where power at certain frequencies can serve as biomarkers that indicate pathological states, time-frequency analysis is required. Frequency data provides information on where the power is concentrated for each pathological state, and the time domain data provides the instant they occur. Short time Fourier transform (STFT) and discrete wavelet transform (DWT) are examined and compared.

1) STFT: It uses the fast Fourier Transform (FFT) to obtain time-frequency data. This is achieved by dividing the signal into windows and FFT is applied to each window [33]. Mathematically STFT is given by

$$X_n\,[t; f] = \sum_{n=0}^{W-1} x[t+n]\ e^{-j\frac{2\pi}{W}nf} \quad (6)$$

where t is the discrete time index, W is the window length into which the signal is split and f is the discrete frequency index. For this application, the time-stamped measurements are split into 2 seconds overlapping epochs, with 50% overlap between epochs. In addition, the power bands (features) are divided into 5 Hz bands, with 3 Hz overlap between bands; 0–5 Hz, 3–8 Hz, … 45–50 Hz. This provides a total of 16 features. The window is chosen such that a balance between time and frequency resolution is obtained.

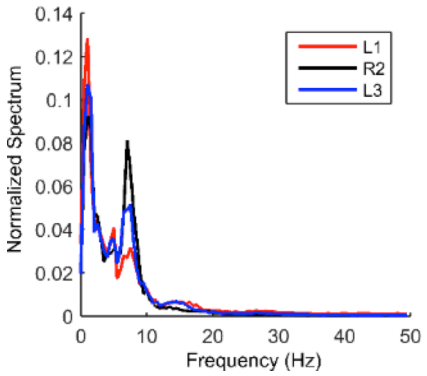

Fig. 3. The power spectrum (PD) of some of the channels in dataset C, shown having nearly similar characteristics. L or R is for Left or Right electrodes respectively, which are numbered from 0 to 3 (caudal to cranial contacts).

2) DWT: It is a time-frequency representation that uses multi-resolution transformation. Mathematically DWT is given by

$$W(u,2^j) = \sum_{n=-\infty}^{\infty} s(n)\,\frac{1}{2^{j/2}}\,\Psi\left(\frac{n-u}{2^j}\right) \qquad (7)$$

where u is the translational parameter representing the time axis, 2^j (j is an integer) is the scale parameter representing the frequency axis and ψ is the wavelet function [34]. Based on the scale parameter in (7), at each level (j), it is down-sampled by 2 to the power of that level (2^j). For the DWT, a 4-level decomposition using the Haar wavelet was obtained. The average power at each level of decomposition consisting of 2 seconds overlapping epochs, with 50% overlap between epochs are obtained as features. The five features are defined as detail coefficient level 1 (32–49.5 Hz), level 2 (16–32 Hz), level 3 (8–16 Hz), level 4 (4–8 Hz) and approximation coefficient level 4 (0.5–4 Hz).

B. Dimensionality Reduction (DR)

DR involves reducing the number of features that will be used for patient state detection. A high number of features increases the possibility of data over-fitting, which results in poor generalization of unseen data. Also, periodically changing the extracted features used results in dynamic FE. This work explores the principal component analysis (PCA), and the MRM (which is eventually chosen for this work).

1) PCA: In PCA, the orthogonal basis (or principal components) that indicates the principal directions in which data varies is calculated [35]. In (8), high dimension features X can be reduced to low dimension features Z,

$$\boldsymbol{Z} = \mathbf{PC_i}^T \times \boldsymbol{X} \qquad (8)$$

where $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ consists of the training features, m is the number of training features and n is the feature vector dimensions. Additionally, $\mathbf{PC} \in \mathbb{R}^{n\times n}$ contains the coefficients of all principal components, $\mathbf{PC_i} \in \mathbb{R}^{n\times i}$ contains the coefficients of the principal component up to the i-th principal component, making i the reduced feature vector dimension and $\boldsymbol{Z} \in \mathbb{R}^{i\times m}$ consists of the new training features with reduced dimensions.

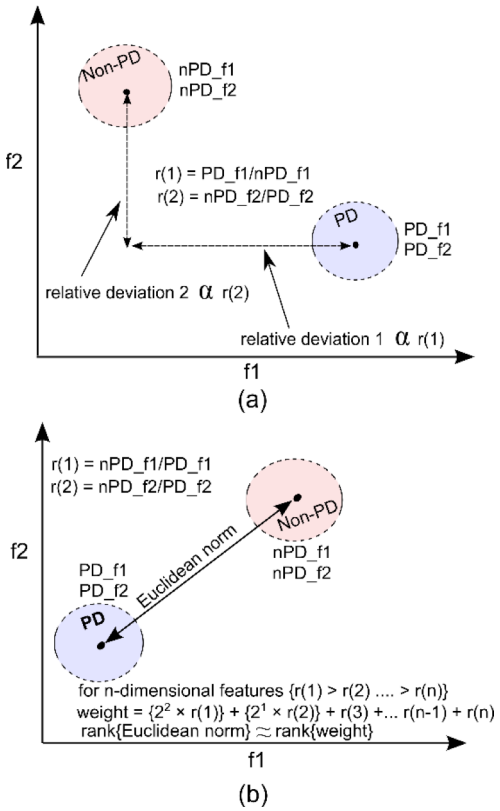

Fig. 4. Two-dimensional depiction of the desired orientation for PD and non-PD clusters in order to create the largest variation using the two features (f1, f2) with the maximum ratio (a) Scenario 1: feature space formed when each of the clusters (PD and non-PD) has the maximum sum for one of the features. In this case the PD cluster has higher f1 values, while the non-PD cluster has higher f2 values. (b) Scenario 2: feature space formed when only one of the clusters has the maximum sum for both features (Non-PD in this case). It also demonstrates how the channel with the highest Euclidean norm between clusters is approximated using weighted ratios.

In addition to feature dimension reduction, DR will be used for channel selection. For PCA, the channel with the least projection error using two dimensional projection is selected. This was done because PD tracking using recordings from all channels may introduce redundancy as recordings from various channels might provide nearly similar information, as can be seen in the spectrum of the channels in Fig. 3 which have nearly similar characteristics. On the other hand, randomly selecting any one of the channels for FE can be counterproductive, as the channel with the least variation between PD and non-PD clusters may be selected. This necessitates the use of a methodical approach for channel and feature selection.

2) MRM: Unlike PCA, the novel MRM is a DR method that uses labeled samples during training. The process of MRM starts by identifying the channel having features with the most pronounced variation in activity. The goal is to obtain the feature space depicted in Fig. 4 which makes classification easier. The MRM is a computationally simple method. Using example values, the steps are described in Fig. 5 and are outlined below:

1. LFP recordings from each monopolar channel are split into M training epochs with equal number of PD and non-PD training examples. In Fig. 5, $M = 128$ results in 64 epochs for both PD and non-PD.
2. Each epoch is divided into N bands (features). For each feature, the sum of that feature for all training examples for both PD and non-PD cases are obtained. In Fig. 5, $N = 5$ is selected with the power in each of the five bands for both PD and non-PD training examples shown.
3. For the summed features (calculated in step 2), the ratio of correspondingly indexed features for PD and non-PD are calculated. The smaller of the two is made the divisor. The division is indicated in red in Fig. 5 with the obtained result shown in step 3 of Fig. 5.
4. The ratios are arranged in descending order. This order shows the relative variation for each feature between PD and non-PD for the training examples – from the largest to the smallest.
5. The maximum ratio is multiplied by 4 and the second maximum ratio is multiplied by 2 and the rest are unchanged. If k features are used for classification, the sum of the first k features gives the channel weight. In Fig. 5, if weights based on two features are required, step 5 shows the channel weight in red which is 38. The channel weight gives a low computation approximation of the channel with the largest Euclidean norm between PD and non-PD clusters for the selected features. The approximate Euclidean norm rank is obtained using a modified version of [36].
6. Steps 1 to 5 are repeated for the rest of the channels. The channel with the largest weight is selected and recordings from that channel are used for classification until another training phase, after which the new highest ranked channel is adopted. Using the values in step 6 of Fig. 5, the channel with the maximum weight is selected, which is channel #0 with a weight of 38.

C. Dynamic Pattern Classification

Dynamic classification uses a modified version of traditional pattern classifiers in order to accurately track the nonlinearities in the extracted features. The classifiers must not be so simple that they are unable to distinguish between classes, yet not so complex as to over-fit the training data [35]. The best classifier will be selected based on a trade-off between computational complexity and performance. The dynamic pattern classifier steps through three orders of the traditional pattern classifiers to be evaluated. The first order classifier is used if it achieves a classification accuracy greater than 90% on validation data, else a higher order classifier is invoked until the criteria is satisfied. If the criteria is not satisfied, the best performing classifier of the three is used. Below is a brief description of the evaluated pattern classifiers.

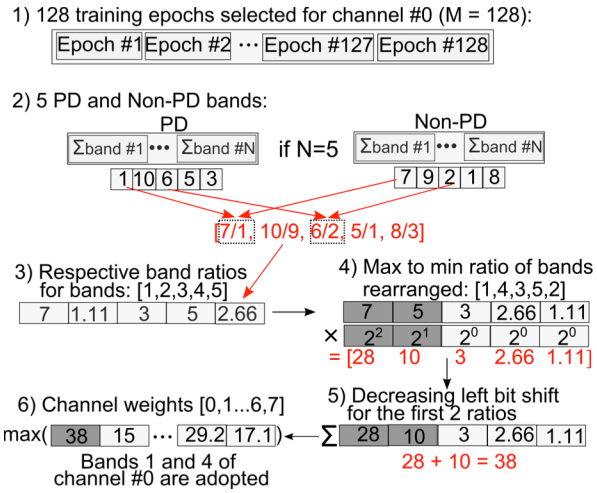

Fig. 5. Detailing the maximum ratio method (MRM). This shows a situation where the channel and feature selection for MRM uses only the two prominent features in ranking the channels.

1) Dynamic k-Nearest Neighbor (k-NN): k-NN uses a non-parametric method for classification. Amongst the various classes to distinguish between, it uses the predominant k closest samples in the feature space in classifying unlabeled points, where k is a natural number [37]. The function $f_{\mathrm{kNN}}(x)$ for k-NN is

$$f_{\mathrm{kNN}}(x) = \sum_{n \in N_k(x)} y_n \tag{9}$$

where x is the test case, y_n are the labels for the training datasets, $N_k(x)$ is the index of k-nearest neighbors of x in the training set. Generally, k-NN does not require the normal iterative learning phase necessary in order to fit the training data to a classification model. The dynamic k-NN steps through 3-NN, 5-NN and 7-NN in that order. Odd nearest neighbors are employed for the dynamic classifier because there are two classes and simple majority voting is required.

2) Dynamic Logistic Regression (LR): LR uses probability of class membership for predicting a test case. Considering a binary classification problem, with class membership $y \in \{0, 1\}$, 1 for PD cases and 0 for non-PD cases, $x^{(i)}$ the extracted features for training examples with corresponding $y^{(i)}$ labels. Then the logistic function $f_{\mathrm{LR}}(x)$ for classification in (10) is [38],

$$f_{\mathrm{LR}}(x) = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \tag{10}$$

which produces an $f_{\mathrm{LR}}(x)$ between 0 and 1, in order to predict the binary states. θ^T is a vector of threshold and weight parameters that is specific to a dataset. Using the logistic function $f_{\mathrm{LR}}(x)$, the two possible outcomes of PD detection can be represented as

$$\mathrm{PD} = \begin{cases} 1 & \text{if } f_{\mathrm{LR}}(x) \geq 0.5 \\ 0 & \text{if } f_{\mathrm{LR}}(x) < 0.5 \end{cases} \tag{11}$$

where true PD detection is signified as 1, and false PD detection as 0. The polynomial order of the function $\theta^T x$

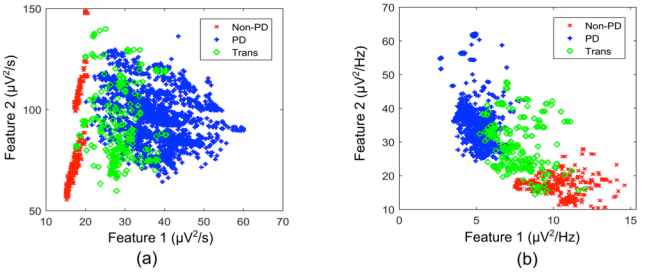

Fig. 6. Feature space formed using MRM for the two prominent features of the selected channel of dataset C using: (a) DWT (b) STFT. LFP epochs at transition points (consisting of PD and non-PD of equal length) are labeled as "Trans".

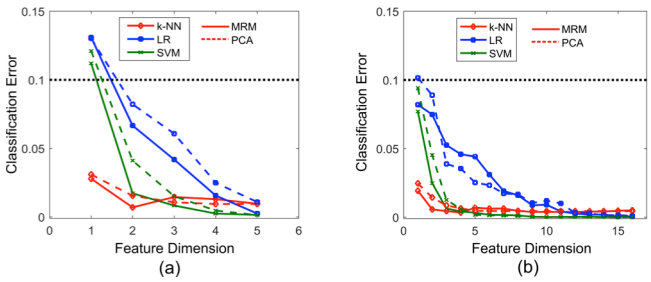

Fig. 7. Effect of feature vector dimension on classification error using: (a) DWT (b) STFT. The plots with the dashed lines are those using PCA for DR, while those without dashes use MRM for DR. The goal is to determine the minimum number of features that achieve a classification accuracy of 90% (10% classification error).

determines the type of LR. For the dynamic LR, steps through a linear function, third order and fifth order classifiers in that order. Functions having only odd degree polynomials are tested since the parameter e in the logistic function $f_{\mathrm{LR}}(x)$ in (10), needs to be raised to a negative power.

3) Dynamic Support Vector Machine (SVM): SVM uses the widest margin between differing states to discriminate. In (12), the discriminating function $f_{\mathrm{SVM}}(x)$, used in classifying test cases is obtained using the training examples as [39]

$$f_{\mathrm{SVM}}(x) = \sum_i y_i \alpha_i K(x_i, x) + b \qquad (12)$$

where x_i are the support vectors and their labels y_i, x is the test case, $K(x_i, x)$ is the kernel transformation, α_i is a weight vector and b represents the classification threshold. In order to handle the complex nature of physiological signals, the decision function can be transformed for use with different kernels, notably the linear kernel, polynomial kernel and the radial basis function (RBF). This can be achieved by replacing the $K(x_i, x)$ in the kernel transformation, with the appropriate kernel function. For dynamic SVM classification, a linear kernel, quadratic kernel and RBF are used in that order.

IV. Results

The results were tested in MATLAB on the described LFP test datasets.

A. Feature Space

A visualization of the feature space formed by the two features selected using MRM for dataset C is shown in Fig. 6. The STFT shows a clearer separation (for PD, Trans and non-PD training examples) compared to the DWT, due to the use of a reduced frequency by the STFT. A reduced band provides better confinement of the relevant frequencies and reduces sensitivity to outliers. The two methods are compared to identify the optimum performance in terms of detection accuracy and complexity. Generally, STFT is ideal for capturing sinusoidal features, and the DWT is ideal for detecting non-continuous frequencies. This was concluded in [34] using the Haar wavelet.

For toolbox based analysis of DWT (such as FieldTrip and EEGlab), Morlet wavelets are used. For toolbox based STFT, spectral smoothing is introduced using multi-tapering. However, time-frequency analysis in this work was guided by considerations for hardware implementation, which are not fundamental for toolbox based analysis. For hardware implementation, the possible improvement in accuracy due to spectral smoothing, is not commensurate with the computational cost incurred. As demonstrated in Section IV.B on accuracy, all the examined combinations achieve less than 10% classification error even without spectral smoothing. Both DWT (Haar wavelet) and STFT (without multi-tapering) have been shown as suitable in hardware-aware implementations for time-frequency analysis [40].

B. Complexity and Accuracy Measures

Various measures such as choice probabilities, classification error and F1-score were used in order to evaluate the detection methods. In each case the average result from 100 runs of Monte Carlo analysis was found in order to obtain the general trend.

1) Error Calculations: For the classification error averaged over all datasets, the goal was to obtain the minimum parameters (minimum number of features, minimum level of quantization and minimum training examples) that resulted in 90% classification accuracy (10% classification error), so as to compute the complexity of each combination. Fig. 7 presents the effect of varying the features used in classification. In Fig. 7(a), the results for the combinations that use DWT for FE are presented; the k-NN based algorithms (combinations) present the best performance, having a classification accuracy greater than 90% irrespective of the number of features used for classification. The second best performance was obtained by the SVM based classifiers, with the DWT-MRM-SVM having a superior performance compared to the DWT-PCA-SVM, even though both require at least two features to attain the 90% accuracy mark. The LR based algorithms behave in a similar way as the SVM based algorithms; however, they have a more gradual slope. In Fig. 7(b), the combinations using STFT for FE have nearly identical characteristics as those using DWT for FE. That is, the k-NN based classifiers have the best performance, followed by the SVM based classifiers and then the LR-based classifiers. Like the DWT algorithms in Fig. 7(a), the STFT algorithms in Fig. 7(b) achieve an accuracy of 90% with two or less features.

With respect to the minimum training examples, Fig. 8(a)

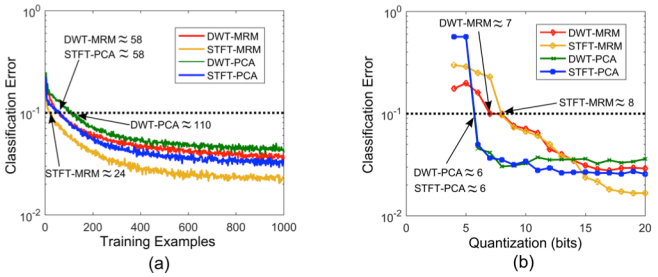

Fig. 8. Optimum values of complexity parameters using one feature classification for k-NN in order to achieve 90% classification accuracy (a) Optimum number of training examples. (b) Quantization (in bits).

presents the performance of the algorithms using k-NN based classifiers. For a classification accuracy of 90%, algorithms using MRM for DR require many fewer training examples: 24 for STFT-MRM-KNN and 58 for DWT-MRM-KNN. However, the PCA based algorithms require at least 58 training examples for STFT-PCA-KNN and 110 for DWT-PCA-KNN. Another notable characteristic is that the combinations using STFT require fewer training examples compared to those using DWT. This may be attributable to the narrower frequency bands used in STFT, which makes it easier to discern patterns with few training examples – wider frequency bands like in DWT may be more susceptible to noise.

For the required bit quantization in Fig. 8(b), the PCA based algorithms outperform their MRM counterparts. The PCA based algorithms requires fewer quantization levels possibly because of the need for less detail in the number representation after PCA transformation. Training examples are less packed together after PCA transformation, because PCA transforms features to a new feature space with a higher variance. The need for less detail after PCA transformation enables the use of fewer quantization levels. The minimum parameters needed in order to achieve a classification accuracy of 90% for the rest of the algorithms is summarized in Table III. It can be seen that the k-NN and LR based algorithms have a higher median number of training examples, compared to the SVM. This is because the k-NN and LR are population dependent algorithms that extrapolate properties more accurately when larger training sets are used. In Table III, it can be seen that 7 combinations require only one feature to achieve 90% classification accuracy. In terms of the quantization, the median is 6 bits resolution (excluding the shaded rows). The shaded rows are the parameters that achieve 90% classification accuracy if all features are used for dynamic detection. From Table III for the MRM algorithms, when the maximum number of features are used, the required training examples and quantization needed to achieve 90% classification accuracy are reduced. However, for the STFT-MRM combinations (STFT-MRM-KNN, STFT-MRM-KNN and STFT-MRM-SVM) using all features presents a lower NOP for 90% classification accuracy compared to its low feature alternative. This may be because computing STFT and extracting power from a single band (single feature) requires slightly less computation than extracting power from all the

TABLE III
COMBINATIONS OF ALGORITHMS AND THE LEVEL OF COMPLEXITY NEEDED TO ACHIEVE 90% CLASSIFICATION ACCURACY

Combination	Number of Features	Quantization (bits)	Number of Training Examples	Complexity	
				NOP ($\times 10^5$)	Area (mm^2)
DWT-MRM-KNN	1	7	58	0.26	0.84
	5	5	14	0.61	1.48
DWT-PCA-KNN	1	6	110	0.24	0.81
STFT-MRM-KNN	1	8	24	0.90	2.12
	16	4	22	0.36	1.15
STFT-PCA-KNN	1	6	58	0.56	1.40
DWT-MRM-LR	2	5	16	0.24	1.05
	5	8	8	1.51	3.98
DWT-PCA-LR	2	6	236	0.39	1.84
STFT-MRM-LR	1	8	28	0.89	2.70
	16	4	20	0.31	1.26
STFT-PCA-LR	2	6	102	0.56	1.96
DWT-MRM-SVM	2	5	14	0.24	1.05
	5	4	8	0.39	1.25
DWT-PCA-SVM	2	6	30	0.37	1.45
STFT-MRM-SVM	1	8	26	0.89	2.70
	16	4	8	0.29	1.11
STFT-PCA-SVM	1	6	170	0.57	1.94

allocated bands (16 in this case). However, the extra computation incurred is offset by the fewer training examples and quantization levels required when all the bands (features) are used.

Mathematically the NOP calculation is given by

$$\text{NOP} = \frac{\text{Training NOP}}{86{,}400} + \text{Operating NOP}. \tag{13}$$

The training NOP is divided by 86400 because it is assumed that training occurs once in a day (86,400 seconds), and the patient state is updated every second during normal operation.

2) Choice Probability and F1-Score: Using the optimum combination for quantization, features and training examples that achieved 90% classification accuracy (summarized in Table III), Fig. 9(a)–9(d) present the performance of the algorithms in terms of their F1-score and CP. For the classifiers, it is clear that the k-NN based algorithms present the best performance (having the least error cost), followed by the SVM-based and then the LR-based. The STFT-MRM-KNN has the least error cost of 0.0075 [in Fig. 9(b)], while the STFT-PCA-SVM has the maximum error cost with 0.0778 [in Fig. 9(d)].

C. Combined Cost

The combined cost represents the cost incurred by a given combination. Fig. 10 shows a plot of error cost vs complexity cost. An ideal detection algorithm is required to have a combined cost at the origin of Fig 10. The error cost is obtained at 10% classification error and consists of costs resulting from a low average choice probability, and a low

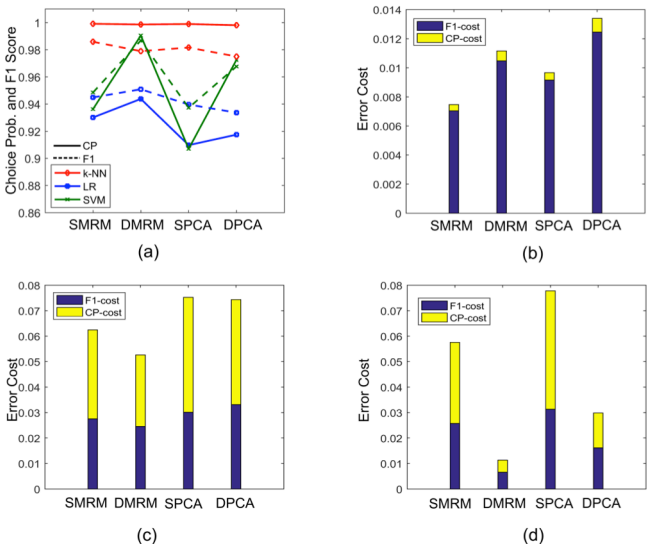

Fig. 9. Plot of choice probability (CP) and F1-score (SMRM is STFT-MRM, DMRM is DWT-MRM, SPCA is STFT-PCA and DPCA is DWT-PCA). (a) Plot of CP and F1-score for different combination of algorithms; (b) Plot of error cost for algorithms using k-NN for classification; (c) Plot of error cost for algorithms using LR for classification; (d) Plot of error cost for algorithms using SVM classification.

average F1-score. The choice probability is included because it shows how well a detector ranks PD cases compared to non-PD cases. The F1-score is included because it shows how good the precision and sensitivity of the detector is. These two factors are not apparent when only classification error is used in assessing the error cost of a detector. They are added to ensure that the error cost is robust enough to cover all accuracy measures. For the complexity cost, the NOP and estimated area are used as the measures. In Fig. 10, the detector with the low costs are indicated by blue markers, and are the DWT-MRM-KNN, DWT-MRM-SVM and DWT-PCA-KNN. There are six medium cost detectors indicated by green markers. The high cost detectors are indicated by red markers and are the STFT-PCA-SVM, DWT-PCA-LR and STFT-PCA-LR. The algorithms using DWT for FE, are mainly closer to the origin, hence the DWT based algorithms have the optimal trade-off between complexity and accuracy. For DR, the MRM based algorithms have the lowest cost while for classification the k-NN based algorithms have the lowest cost. Thus, the combination closest to the origin is the DWT-MRM-KNN, which represents the optimal trade-off between accuracy and complexity; while the worst performing algorithm is the STFT-PCA- SVM.

V. Discussion

A. Semi-Synthetic Datasets

Semi-synthetic data was generated by taking advantage of statistical properties in the data to fit ARMA models. Moreover, for every single epoch a range of Monte Carlo forecasts were simulated based on the underlying variations. To the best of our knowledge, there is no literature on semi-synthetic LFP data generation using real-life recordings. However, for other signals such as spikes and ECG there is

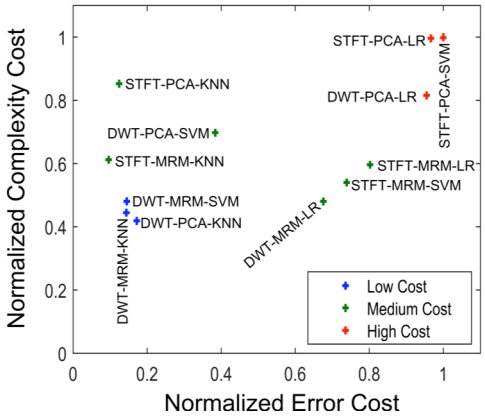

Fig. 10. The combined cost for different combinations of algorithms. Both axes are normalized such that the maximum cost in each case is equal to 1.

extensive literature. Using statistical parameters like the ones used in defining spike [41] or ECG [42] activity could be misleading; because they have a unique morphology which can be varied using simple measures, such as the amplitude or shape. On the other hand, LFPs result from the activity of a localized population of sources; hence using statistics that are not population based corrupts and destroys the signal fidelity. In this work a number of properties were exploited, among which are:

1) Statistical Similarity: To measure the sensitivity of the algorithms to all forms of PD and non-PD variations, it was ensured that statistically dissimilar epochs (measured using the normalized cross-correlation) were attached together during periods of long PD or long non-PD synthesis. This introduced some randomness by attaching epochs with varying similarity, which enabled the assessment of the algorithms' robustness to instantaneous changes within the same patient state.

2) Dispersion and Random Permutation Sampling: Dispersion was introduced using ARMA models to forecast a range of Monte Carlo variants for each epoch. This was to create a large diversity pool that overcomes sampling bias (since sampling bias can lead to poor generalizability). Epoch selection using random permutation sampling was applied to avoid data leakage which can cause over generalizability.

3) Poisson Distribution Defined PD and non-PD Duration: Signal length was defined using a Poisson distribution to make PD and non-PD episodes pseudo-random. This ensured that a "randomly guessing" algorithm that changes state prediction based on a predefined pattern is flagged because of the pseudo-random PD and non-PD periods.

B. Spectral Bands and MRM

In the normalized autospectra shown in Fig. 11 the largest variation between the PD and non-PD autospectra is between 10–25 Hz, which mostly lies in the beta band; 13–30 Hz. The gamma band (> 30 Hz) shows little relative activity. However, Table IV in which the two frequency bands with the most activity for each of the nine datasets is summarized, it can be seen that the gamma band is not strictly without activity. Datasets A, F and G have their most pronounced

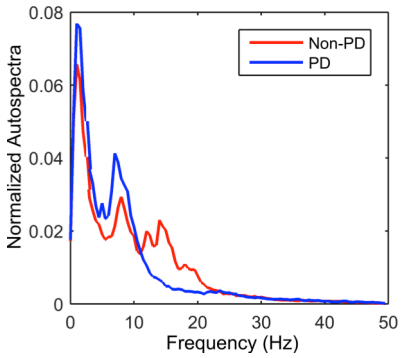

Fig. 11. Average normalized autospectra for all the datasets. The PD and non-PD plots are normalized such that the total autospectra sum in each case is equal to 1.

TABLE IV
SUMMARIZING THE TWO FREQUENCY BANDS WITH THE MOST PRONOUNCED VARIATION

Dataset	STFT		DWT	
	Maximum Variation Band (Hz)	2nd Maximum Variation Band (Hz)	Maximum Variation Band (Hz)	2nd Maximum Variation Band (Hz)
A	6–11	9–14	32–49.5	4–8
B	12–17	15–20	16–32	8–16
C	21–26	18–23	32–49.5	16–32
D	0.5–5	6–11	4–8	8–16
E	12–17	15–20	8–16	16–32
F	18–23	15–20	32-49.5	16–32
G	45–49.5	42–47	32–49.5	16–32
H	36–41	33–38	0.5–4	4–8
I	18–23	21–26	16–32	32–49.5

variation in the gamma bands (level 1 detail coefficients) when DWT is used for FE. While using STFT, it was only visible for dataset G. The little activity in gamma bands is corroborated by [13], in which there are occasions where there is little activity, while at other times they have the most pronounced activity.

The frequency bands for the STFT and DWT in Table IV appear to be unexpectedly different. This is due to different sized frequency bands. For instance, in dataset (patient) D, the 0.5–5 Hz band presents the most pronounced variation for STFT, while the 4– Hz band provides the most variation for DWT. This shows that most of the power is within the 4–5 Hz band, giving rise to the dominance of the bands containing these frequencies. The difference between the STFT and DWT in other cases can be explained because the most pronounced variation is dependent on the relative power between PD and non-PD. That is, there are cases in which a PD band can have the greatest absolute activity compared to other PD bands. But when compared to its corresponding non-PD band, it may not have the most pronounced variation. This was the case for datasets A and H. As a result of the varying frequency spread present, particularly in the DWT, only the activity of the much lower bands of the DWT should be expected to show some semblance in behavior to the STFT. These are the cases were the frequency spread is quite close and at the higher

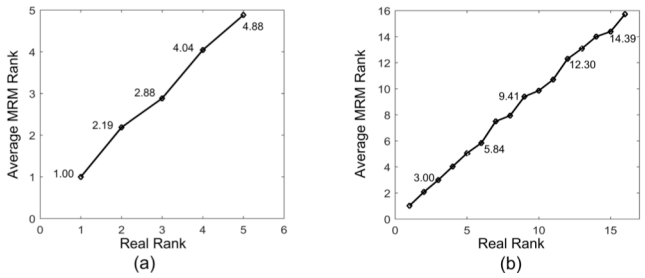

Fig. 12. Shows the mean plot of how the MRM algorithm ranks the frequency bands using: (a) DWT for FE (b) STFT for FE. The rank is from the band with the most pronounced variation (ranked 1) to the least pronounced (ranked 5 for DWT and 16 for STFT).

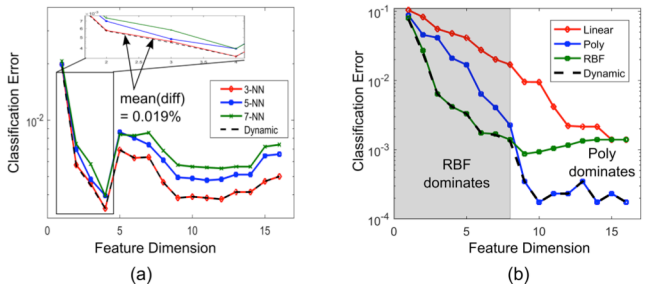

Fig. 13. Comparison of the classification error for static and dynamic detection schemes: (a) using the STFT-MRM-KNN detector; (b) using the STFT-MRM-SVM detector.

frequencies, they cannot be compared because a single DWT band covers more than five STFT bands.

In Fig. 12, it is shown that the MRM algorithm can detect the bands with the most pronounced variation even though the training examples are a small fraction of the entire population. In Fig. 12(a), it accurately detects the band with the most pronounced variation for all 100 runs while in Fig. 12(b) it accurately detects the three prominent bands for all runs. Most of the average ranks approximate to the real rank showing that most of the time, it ranks the bands in their right order. This demonstrates that the training examples used have enough diversity present in the entire signal. The MRM algorithm also assists in channel selection by approximately determining the channel with the largest Euclidean norm between PD and non-PD clusters. Hence, instead of using all eight monopolar channels for detection, only the channel with the most pronounced variation is selected. This makes classification less computationally intensive.

C. Dynamic Detection

The dynamic schemes are used to obtain a classifier that is compatible with the dynamic FE and at the same time achieves the right balance between complexity and classification accuracy. The dynamic classifier operates in two phases – concurrent detection and training, and detection only. Compatibility between dynamic FE and dynamic pattern classification is determined periodically during the concurrent detection and training phase. The dynamic detection operates in real-time since both phases involve detection. Determining

compatibility is necessary because high order classifiers may sometimes provide inferior performance compared to low order classifiers, making the use of high order classifiers redundant. A classic example of this is shown in Fig. 13(a), which uses the STFT-MRM-KNN algorithm. Apart from the region where few features (one to four features) are used in which there is a struggle for dominance, the 3-NN completely dominates when five or more features are used. Thus, making the higher order classifiers (5-NN and 7-NN) redundant. Nevertheless, the benefits of dynamic detection are obvious since a mean reduction of 0.019% in classification error is obtained between the dynamic classifier and the 3-NN. This is because different kernels are selected for different datasets (A–I). In the long term, these slight differences of 0.019% between the dynamic classifier and the best performing static classifier can cumulatively lead to more beneficial effects in therapy since there is a higher likelihood for detection. Conversely, in Fig. 13(b), which depicts the result for the STFT-MRM-SVM algorithm, there are two regimes of operation: when 8 features or less are used for classification, the RBF, which is the highest order kernel dominates. On the other hand, when more than 8 features are used the lower order poly kernel dominates, similar to Fig 13(a). This shows that dynamic detection takes advantage of the best operating regimes of various kernels to produce performance superior to those of static classifiers as is shown in Fig. 13. As a result the three different classifiers in the dynamic scheme are complementary.

VI. Conclusion

Dynamic PD detection is beneficial for on-demand DBS because it personalizes PD detection. This eases tracking of the dynamic variations common in PD pathophysiology. The results presented show that on-chip (online) PD detection is possible. It has been demonstrated that the change in power characteristics of LFP signals, can be tracked using a combination of DWT for FE, dimensionality reduction using the MRM (a novel algorithm proposed in this paper) and dynamic k-NN for classification. This combination can create an efficient algorithm that has the best trade-off between computational complexity and detection accuracy. In addition to PD detection, the algorithm could be extended to other applications that require on-demand DBS for efficient modulation of therapy.

Acknowledgement

The authors would like to thank Prof. Peter Brown at the University of Oxford for providing the clinical recordings.

References

[1] S. Little, A. Pogosyan, S. Neal, B. Zavala, L. Zrinzo, M. Hariz, T. Foltynie, P. Limousin, K. Ashkan, J. Fitzgerald, A. L. Green, T. Z. Aziz, and P. Brown, “Adaptive deep brain stimulation in advanced Parkinson disease,” *Ann. Neurol.*, vol. 74, no. 3, pp. 449–457, Sep. 2013.

[2] A. Priori, G. Foffani, L. Rossi, and S. Marceglia, “Adaptive deep brain stimulation (aDBS) controlled by local field potential oscillations,” *Exp. Neurol.*, vol. 245, pp. 77–86, Jul. 2013.

[3] I. Basu, D. Graupe, D. Tuninetti, P. Shukla, K. V Slavin, L. V. Metman, and D. M. Corcos, “Pathological tremor prediction using surface electromyogram and acceleration: potential use in ‘ON-OFF’ demand driven deep brain stimulator design,” *J. Neural Eng.*, vol. 10, no. 3, p. 036019, Jun. 2013.

[4] M. Kilpatrick, E. Church, S. Danish, M. Stiefel, J. Jaggi, C. Halpern, M. Kerr, E. Maloney, M. Robinson, I. Lucki, E. Krizman-Grenda, and G. Baltuch, “Intracerebral microdialysis during deep brain stimulation surgery,” *J. Neurosci. Methods*, vol. 190, no. 1, pp. 106–111, Jun. 2010.

[5] M. Arlotti, M. Rosa, S. Marceglia, S. Barbieri, and A. Priori, “The adaptive deep brain stimulation challenge,” *Parkinsonism Relat. Disord.*, vol. 28, pp. 12–7, Jul. 2016.

[6] T. Endo, R. Okuno, M. Yokoe, K. Akazawa, and S. Sakoda, “A novel method for systematic analysis of rigidity in Parkinson’s disease,” *Mov. Disord.*, vol. 24, no. 15, pp. 2218–24, Nov. 2009.

[7] S. Little and P. Brown, “What brain signals are suitable for feedback control of deep brain stimulation in Parkinson’s disease?,” *Ann. N. Y. Acad. Sci.*, vol. 1265, no. 1, pp. 9–24, Aug. 2012.

[8] S. Stanslaski, P. Afshar, P. Cong, J. Giftakis, P. Stypulkowski, D. Carlson, D. Linde, D. Ullestad, A.-T. Avestruz, and T. Denison, “Design and validation of a fully implantable, chronic, closed-loop neuromodulation device with concurrent sensing and stimulation,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 20, no. 4, pp. 410–21, Jul. 2012.

[9] M. Rivlin-Etzion, O. Marmor, G. Heimer, A. Raz, A. Nini, and H. Bergman, “Basal ganglia oscillations and pathophysiology of movement disorders,” *Curr. Opin. Neurobiol.*, vol. 16, no. 6, pp. 629–37, Dec. 2006.

[10] P. F. Grant and M. M. Lowery, “Simulation of cortico-basal ganglia oscillations and their suppression by closed loop deep brain stimulation,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 4, pp. 584–94, Jul. 2013.

[11] S. Little, A. Pogosyan, A. A. Kuhn, and P. Brown, “β band stability over time correlates with Parkinsonian rigidity and bradykinesia,” *Exp. Neurol.*, vol. 236, no. 2, pp. 383–8, Aug. 2012.

[12] M. Alegre and M. Valencia, “Oscillatory activity in the human basal ganglia: more than just beta, more than just Parkinson’s disease,” *Exp. Neurol.*, vol. 248, pp. 183–6, Oct. 2013.

[13] J.-S. Brittain and P. Brown, “Oscillations and the basal ganglia: motor control and beyond,” *Neuroimage*, vol. 85 Pt 2, pp. 637–47, Jan. 2014.

[14] P. Brown and D. Williams, “Basal ganglia local field potential activity: Character and functional significance in the human,” *Clin. Neurophysiol.*, vol. 116, no. 11, pp. 2510–2519, 2005.

[15] T. Heida, E. C. Wentink, and E. Marani, “Power spectral density analysis of physiological, rest and action tremor in Parkinson’s disease patients treated with deep brain stimulation,” *J. Neuroeng. Rehabil.*, vol. 10, no. 1, p. 70, Jan. 2013.

[16] J. F. Baizabal-Carvallo and J. Jankovic, “Movement disorders induced by deep brain stimulation,” *Parkinsonism Relat. Disord.*, vol. 25, pp. 1–9, Jan. 2016.

[17] M. L. Hacker, J. Tonascia, M. Turchan, A. Currie, L. Heusinkveld, P. E. Konrad, T. L. Davis, J. S. Neimat, F. T. Phibbs, P. Hedera, L. Wang, Y. Shi, D. M. Shade, A. L. Sternberg, L. T. Drye, and D. Charles, “Deep brain stimulation may reduce the relative risk of clinically important worsening in early stage Parkinson’s disease,” *Parkinsonism Relat. Disord.*, vol. 21, no. 10, pp. 1177–83, Oct. 2015.

[18] A. Mohammed, M. Zamani, R. Bayford, and A. Demosthenous, “Patient specific Parkinson’s disease detection for adaptive deep brain stimulation,” in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 1528–1531.

[19] S. Santaniello, G. Fiengo, L. Glielmo, and W. M. Grill, “Closed-loop control of deep brain stimulation: a simulation study,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 1, pp. 15–24, Fe. 2011.

[20] S. Marceglia, M. Fumagalli, and A. Priori, “What neurophysiological recordings tell us about cognitive and behavioral functions of the human subthalamic nucleus,” *Expert Rev. Neurother.*, vol. 11, no. 1, pp. 139–49, Jan. 2011.

[21] R. Katzenschlager and A. J. Lees, “Treatment of Parkinson’s disease: levodopa as the first choice,” *J. Neurol.*, vol. 249 Suppl, pp. II19–24, Sep. 2002.

[22] N. Singh, V. Pillay, and Y. E. Choonara, “Advances in the treatment of Parkinson’s disease,” *Prog. Neurobiol.*, vol. 81, no. 1, pp. 29–44, Jan. 2007.

[23] G. Birot, L. Albera, F. Wendling, and I. Merlet, "Localization of extended brain sources from EEG/MEG: The ExSo-MUSIC approach," *Neuroimage*, vol. 56, no. 1, pp. 102–113, May 2011.
[24] S. Makeig, C. Kothe, T. Mullen, N. Bigdely-Shamlo, Zhilin Zhang, and K. Kreutz-Delgado, "Evolving Signal Processing for Brain–Computer Interfaces," *Proc. IEEE*, vol. 100, no. Special Centennial Issue, pp. 1567–1584, May 2012.
[25] J. Muthuswamy, "Biomedical Signal Analysis," in *Standard Handbook of Biomedical Engineering and Design*, M. Kutz, Ed. The McGraw-Hill Companies, 2003, pp. 1 – 30.
[26] C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th ed. Boca Raton, Florida: Chapman & Hall/CRC, 2004.
[27] G. Reinoso, J. C. Allen, W.-L. Au, S.-H. Seah, K.-Y. Tay, and L. C. S. Tan, "Clinical evolution of Parkinson's disease and prognostic factors affecting motor progression: 9-year follow-up study," *Eur. J. Neurol.*, vol. 22, no. 3, pp. 457–63, Mar. 2015.
[28] C. Ding, G. Ganesvaran, J. E. Alty, B. G. Clissold, C. D. McColl, K. A. Reardon, M. Schiff, V. Srikanth, and P. A. Kempster, "Study of levodopa response in Parkinson's disease: Observations on rates of motor progression," *Mov. Disord.*, vol. 31, no. 4, pp. 589–592, Apr. 2016.
[29] J. Semmlow, *Signals and Systems for Bioengineers*. Elsevier, 2012.
[30] S. Gibson, J. W. Judy, and D. Marković, "Technology-aware algorithm design for neural spike detection, feature extraction, and dimensionality reduction," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 5, pp. 469–478, Oct. 2010.
[31] Q. Zou, S. Xie, Z. Lin, M. Wu, and Y. Ju, "Finding the best classification threshold in imbalanced classification," *Big Data Res.*, vol. 5, pp. 2–8, Sep. 2016.
[32] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
[33] P. Prandoni and M. Vetterli, *Signal processing for Communications*. EPFL Press, 2008.
[34] D. Sundararajan, *Discrete Wavelet Transform: A Signal Processing Approach*. New York, United States: John Wiley and Sons Ltd, 2015.
[35] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, Inc., 2001.
[36] M. Barni, F. Buti, F. Bartolini, and V. Cappellini, "A quasi-Euclidean norm to speed up vector median filtering," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1704–9, Jan. 2000.
[37] S. Marshland, *Machine Learning : an algorithmic perspective*, 2nd ed. Boca Raton, FL: CRC Press, 2015.
[38] D. Barber, *Bayesian Reasoning and Machine Learning*. Cambridge, UK: Cambridge University Press, 2012.
[39] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, UK: Cambridge University Press, 2000.
[40] D. Marković and R. Brodersen, *DSP Architecture Design Essentials*. Springer, 2012.
[41] M. Zamani and A. Demosthenous, "Feature extraction using extrema sampling of discrete derivatives for spike sorting in implantable upper-limb neural prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 22, no. 4, pp. 716–726, Jul. 2014.
[42] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 289–294, Mar. 2003.

Ameer Mohammed (S'15) was born in Lagos, Nigeria, in 1989. He completed his undergraduate degree in electrical engineering at Bayero University Kano, Nigeria, graduating top of his class in 2011. He received the M.Sc. degree in Mechatronics from the University of Bath, U.K., in 2013. He then joined the Analog and Biomedical Electronics Group in the Department of Electronic and Electrical Engineering at University College London (UCL), London, U.K., to study for a Ph.D. degree. His research interests are bio-signal processing for implantable devices and low-power integrated circuit design. Mr Mohammed was awarded the Presidential Scholarship from the Federal Republic of Nigeria, to pursue his Ph.D. degree.

Majid Zamani (S'13, M17) was born in Tehran, Iran, in 1984. He received the M.Sc. degree in microelectronics from the Islamic Azad University, Science and Research Branch, Tehran, Iran, in 2011, and was awarded the best researcher M.Sc. student award. He then joined the Analog and Biomedical Electronics Group in the Department of Electronic and Electrical Engineering at University College London (UCL), London, U.K., to study for a Ph.D. degree, which he obtained in 2017. He is currently a Research Associate at UCL. His research interests are the design of implantable devices for neural prostheses, low-power mixed-signal integrated circuits, and data converters. Mr. Zamani was awarded a UCL Overseas Research Scholarship and a Graduate Research Scholarship to pursue his Ph.D. degree.

Richard Bayford (M'85) received the M.Sc. degree in industrial systems from Cranfield University, Cranfield, Bedforshire, U.K., and the Ph.D. degree from Middlesex University, London, U.K., in 1994. He was appointed Professor of bio-modeling and informatics in 2005 in the Department of Natural Sciences, Middlesex University, U.K., and also holds an honorary post in the Department of Electrical and Electronic Engineering, University College London and is a fellow of the Institute of Physics His expertise is in biomedical image/signal processing, electrical impedance tomography, deep brain stimulation, bio-modeling, tele-medical systems, and VLSI design. He has authored over 300 papers in journals and international conference proceedings.

Dr Bayford was the Editor-in-Chief of Physiological Measurement, Institute of Physics (IOP), between 2008 and 2013. He is on the editorial board of the International Journal of Biomedical Imaging, and is the Chair of the journals committee of the Institute of Physics and Engineering in Medicine (IPEM).

Andreas Demosthenous (S'94–M'99–SM'05) received the B.Eng. degree in electrical and electronic engineering from the University of Leicester, Leicester, U.K., the M.Sc. degree in telecommunications technology from Aston University, Birmingham, U.K., and the Ph.D. degree in electronic and electrical engineering from University College London (UCL), London, U.K., in 1992, 1994, and 1998, respectively. He is a Professor in the UCL Department of Electronic and Electrical Engineering where he leads the Analog and Biomedical Electronics Group. He has published over 250 articles in journals and international conference proceedings. His research interests include analog and mixed-signal integrated circuits for biomedical, sensor, and signal-processing applications.

Dr Demosthenous is the Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: REGULAR PAPERS and an Associate Editor for the IEEE TRANSACTIONS ON BIOMEDICAL CIRCUITS AND SYSTEMS. He is a member of the Technical Programme Committee of several IEEE conferences including ESSCIRC and ISCAS. He is a Fellow of the Institution of Engineering and Technology and a Chartered Engineer.